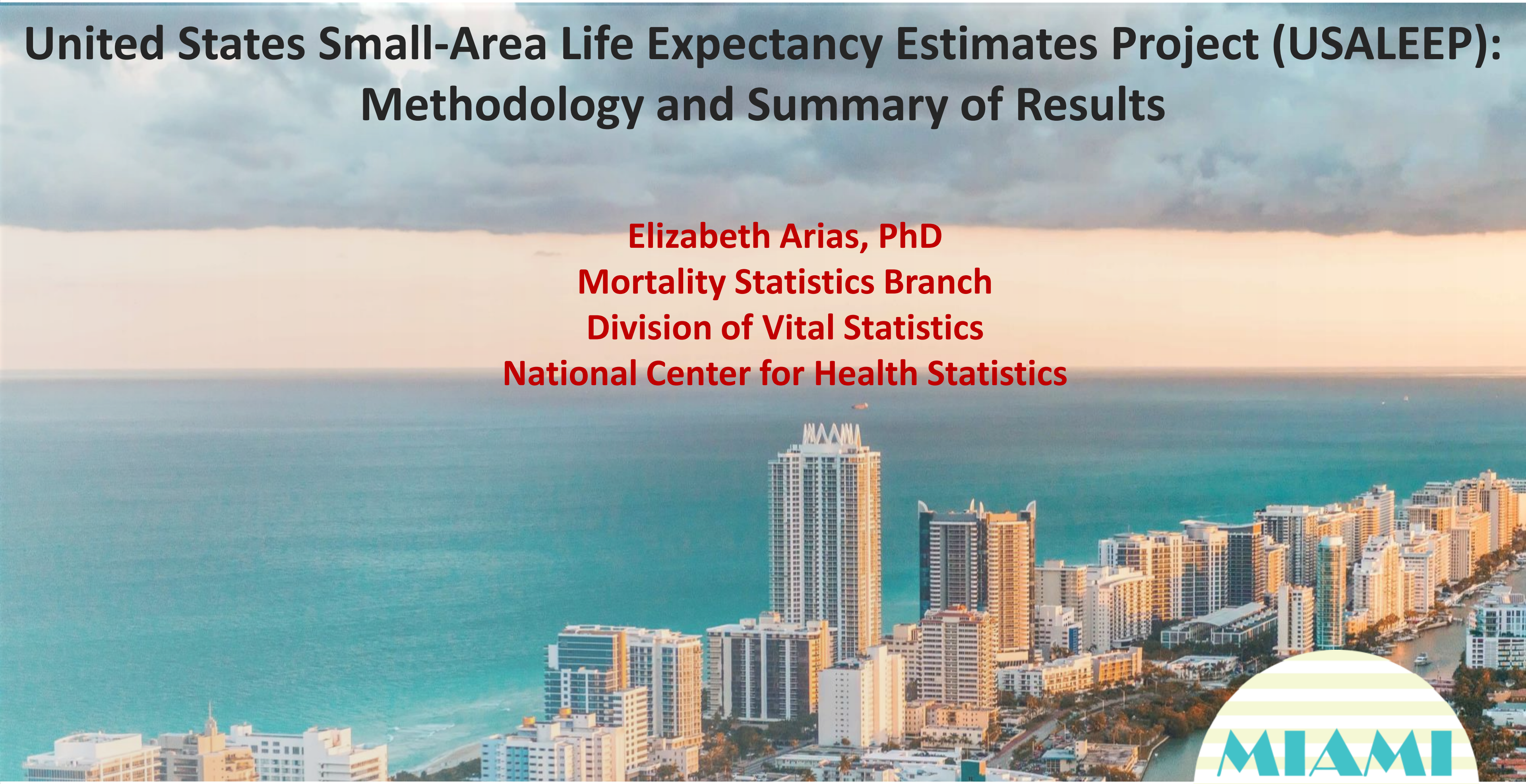
United States Small-Area Life Expectancy Estimates Project (USALEEP): Methodology and Summary of Results

Elizabeth Arias, PhD
Mortality Statistics Branch
Division of Vital Statistics
National Center for Health Statistics



NAPHSIS
Protecting Personal Identity
Promoting Public Health

MIAMI VITALS: THE HEAT IS ON

NAPHSIS ANNUAL MEETING | MIAMI | JUNE 4-7,2018

MIAMI VITALS The Heat is On!

Project Background

- **Goal:** produce estimates of life expectancy at birth for every Census tract in the United States.
- A growing body of research is recognizing the importance of measuring mortality outcomes at the local level to the identification of health disparities within a population
- The indicator most widely identified as the ideal measure of a population's mortality experience is life expectancy at birth.
- Its estimation is complex. It requires the calculation of a life table that includes six distinct functions, a minimum number of age groups, and total population size below which the estimates become unstable and unreliable.

Data and Methodological Challenges

- Mortality data for small areas
  - In the United States, the smallest geographic identifier available nationally for mortality data is the county of residence of decedents.
- Population data for small areas
  - Census Bureau Population Division does not produce postcensal or intercensal population estimates below county level
- Small population sizes and death counts
  - Below a certain size/number, life table estimates are unreliable and unstable
  - How small can populations or death counts be in order to produce reliable life expectancy estimates?
    - Previous studies have convincingly shown minimum population size is 5,000
    - There is no consensus on an acceptable minimum number of deaths

USALEEP Data to Meet Challenges

- **Mortality data**
  - Fifty-three registration areas, all 50 U.S. states, New York City, Washington, D.C. and Puerto Rico, provided to NCHS residential address information for all deaths that occurred in their jurisdictions in years 2010 through 2015.
  - Through a collaborative agreement between NCHS and the Department of Housing and Urban Development (HUD), residential addresses were geocoded by HUD Geocode Service Center (GSC) to identify their corresponding census tract codes. Puerto Rico addresses could not be successfully geocoded and were excluded.
- **Population data**
  - Based on the years of geocoded mortality data, population estimates required would be the mid-period year or an average of mid-year population estimates for the years of interest.
  - The only available sources of census-tract level population data for the years of geocoded mortality data are the ACS 2010-2014 and 2011-2015 5-year surveys.
  - For one of those years, 2010, the 2010 decennial Census April 1 population count is available.
  - Combine the 2010 Decennial Census with the 2011-2015 ACS.   Use 1/6 of the 2010 Census and 5/6 of the ACS. Reduces the variance of the population estimates by a 1/6 because the 2010 Census is a complete count, not a sample as the ACS.

USALEEP Data to Meet Challenges

- The strategy to combine the decennial census counts with the ACS survey-based estimates was the result of an exploration of the benefits and drawbacks of pooling 5 or 6 years of data.

- Pooling over a shorter time span increases the accuracy of the life expectancy estimates. However, the reliability of these diminish when death counts are too small or missing and population estimates have sampling error.

- Adding an additional year of data increased the number of deaths while simultaneously reducing the variance of the population estimates since the decennial census counts are not affected by sampling error; only the ACS 5-year (2011-2015) population estimates are affected by sampling error. As a result, the six-year population estimate is only affected by 5-years of sampling error.

Methodological Strategy

- **Phase 1:**
  - Estimate age-specific death rates, M(x), for tracts with minimum populations of 5,000 (aggregate of 6 years of mortality data) and no missing age-specific death counts.
  - Select tracts with acceptable age patterns, or schedules, of mortality to use as "models" and borrow information for Phase 2 tracts with missing age-specific death counts.

- **Phase 2:**
  - Estimate zero-truncated Poisson or Negative Binomial Models to generate predicted M(x) for Phase 1 "model" tracts and extrapolate to Phase 2 tracts.
  - For Phase 2 tracts with a minimum population size of 5,000 but with missing age-specific death counts, replace missing M(x) with predicted values.
  - Evaluate and select Phase 2 tracts with acceptable age patterns, or schedules, of mortality.

- **All Tracts:**
  - Calculate abridged life tables for all tracts with complete age-specific mortality information and acceptable mortality schedules.
  - Evaluate resulting life tables: explore variance of life table functions, evaluate outliers, compare to US estimates, compare to state-level estimates.

Phase I Tracts

- 5,741 Census Tracts with minimum population size of 5,000 and no zero age-specific death counts
- 4,639 (80.4%) had age-specific mortality patterns that were consistent with the standard age-specific mortality schedule
- Age Pattern of Mortality
  - Pattern should be consistent with the universal age-pattern of mortality; developed rules for identifying acceptable patterns

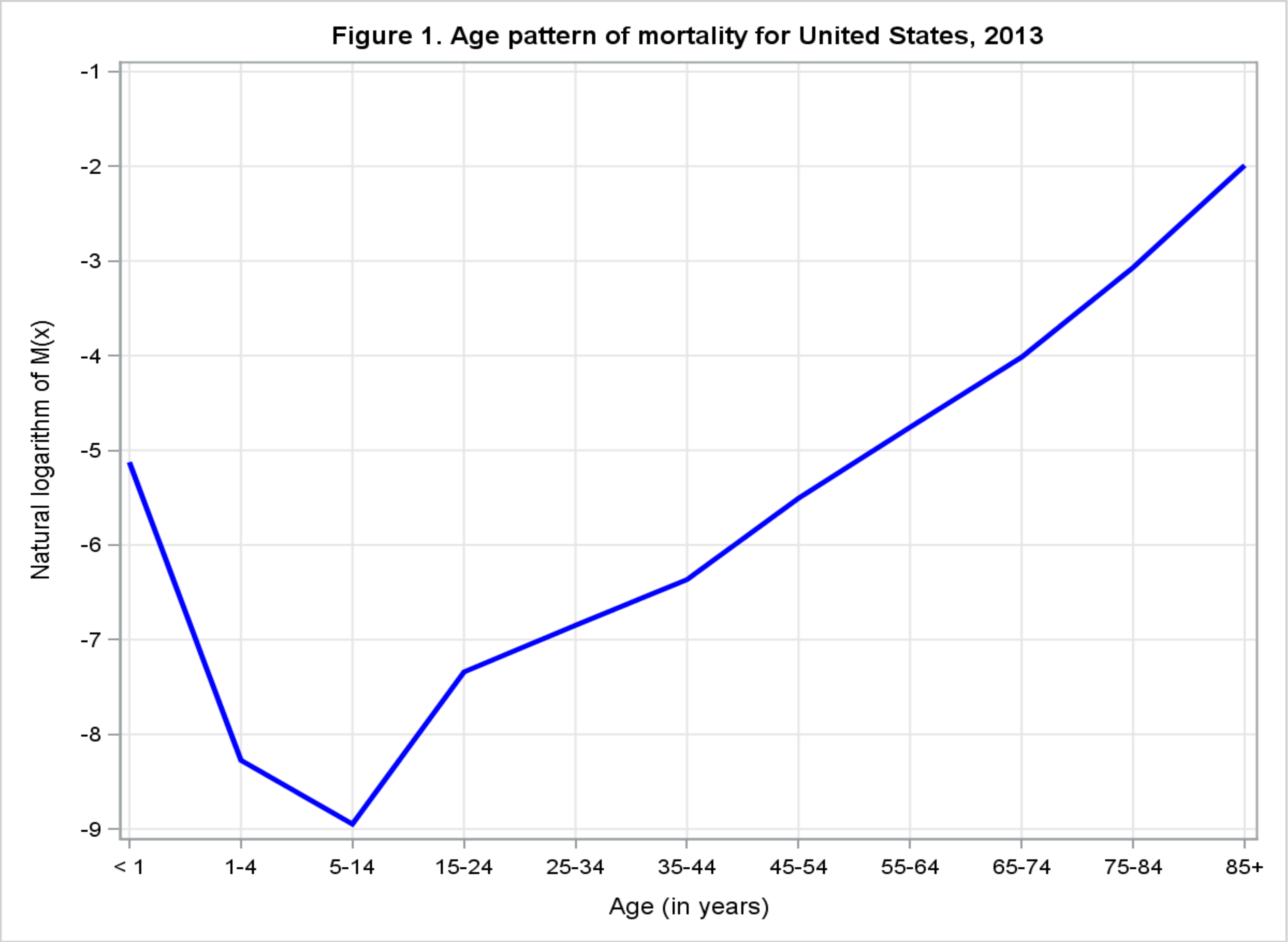


Figure 1. Age pattern of mortality for United States, 2013
Natural logarithm of M(x)
-1
-2
-3
-4
-5
-6
-7
-8
-9
< 1
1-4
5-14
15-24
25-34
35-44
45-54
55-64
65-74
75-84
85+
Age (in years)

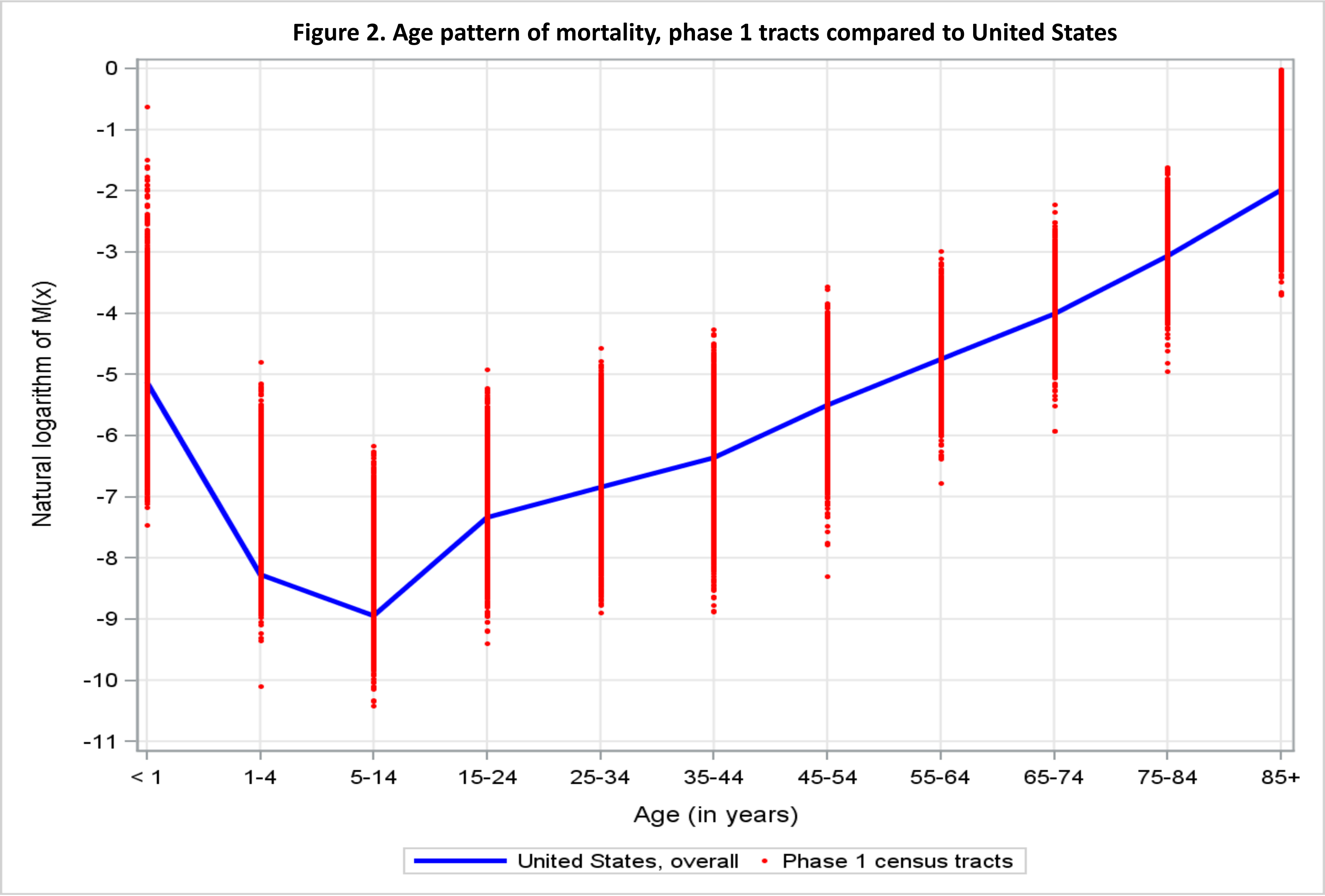


Figure 2. Age pattern of mortality, phase 1 tracts compared to United States
0
-1
-2
-3
-4
-5
-6
-7
-8
-9
-10
-11
Natural logarithm of M(x)
< 1
1-4
5-14
15-24
25-34
35-44
45-54
55-64
65-74
75-84
85+
Age (in years)
United States, overall
Phase 1 census tracts

Statistical Models

- Under the assumption that deaths are Poisson distributed, for each age group we tested the fit of zero-truncated Poisson and Negative Binomial models
- Poisson and negative binomial distributions are the same, but the Poisson assumes the variance is equal to the mean, while the Negative Binomial does not make that assumption, allowing for over dispersion.
- Statistical tests showed there was over dispersion in all age groups. Truncated Negative Binomial models:
  - $_{n}\overline{D}_{x} = e^{\ln({}_{n}P_{x}) + \beta_0 + \beta_1 Y_1 + \cdots + \beta_k Y_k}$
    - $_{n}\overline{M}_{x} = e^{\beta_0 + \beta_1 Y_1 + \cdots + \beta_k Y_k}$

Statistical Models

- Included in models are the following covariates known/hypothesized to be related to mortality:
  - % Hispanic population in the tract
  - % non-Hispanic black population
  - % population with Bachelor’s degree or higher
  - median household income
  - Population density
  - Is tract in a Purchased/Referred Care Area (PRCSDA) IHS health care provision county
  - US region of residence

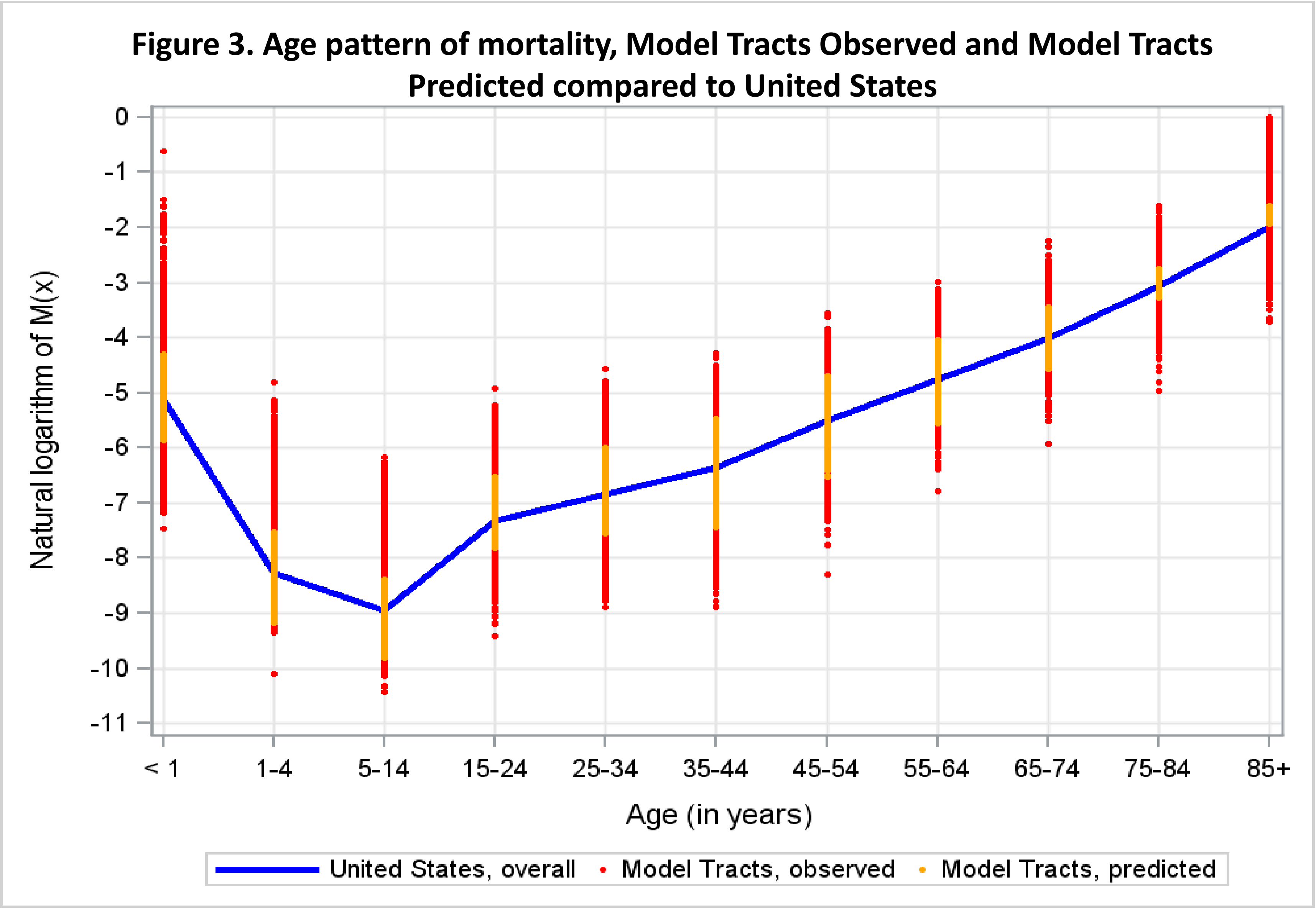
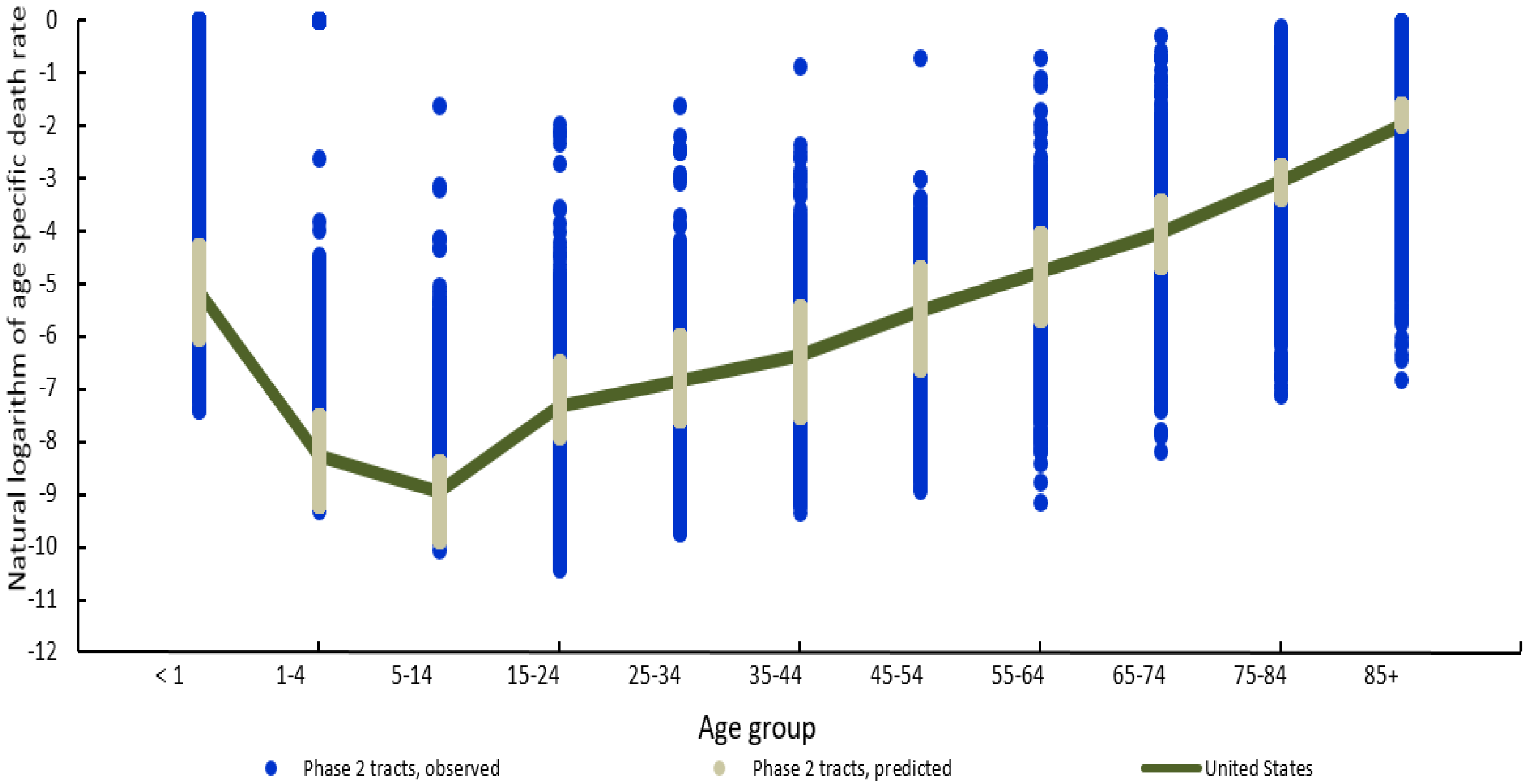


Figure 3. Age pattern of mortality, Model Tracts Observed and Model Tracts Predicted compared to United States
Natural logarithm of M(x)
0
-1
-2
-3
-4
-5
-6
-7
-8
-9
-10
-11
< 1
1-4
5-14
15-24
25-34
35-44
45-54
55-64
65-74
75-84
85+
Age (in years)
United States, overall
Model Tracts, observed
Model Tracts, predicted

Distribution of Phase 2 Census Tracts by Number of Age Categories **with Zero** Deaths and Distribution of Zero Deaths by Age

		Number	Percent			Percent		
Number of Age Categories					Age Categories			
1		17,400	28.70		0-1	13.60		
2		23,643	39.00		1-4	37.00		
3		12,942	21.30		5-14	33.90		
4		4,911	8.10		15-24	8.60		
5		1,398	2.30		25-34	4.40		
6		276	0.45		35-44	1.98		
7		44	0.07		45-54	0.20		
8		12	0.02		55-64	0.07		
9		15	0.02		65-74	0.06		
10		16	0.03		75-84	0.08		
11		28	0.05		85 and Older	0.10		
Total		60,685						

Figure 4. Age patterns of mortality of phase 2 tracts with observed and predicted age-specific death rates, compared to the United States, 2013 schedule



NOTES: Census tracts from Maine and Wisconsin were excluded. Values for the United States, 2013 are based on official statistics published in: https://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_02.pdf

SOURCE: NCHS, National Vital Statistics System, Mortality.

Methodological Challenge of the Life Table

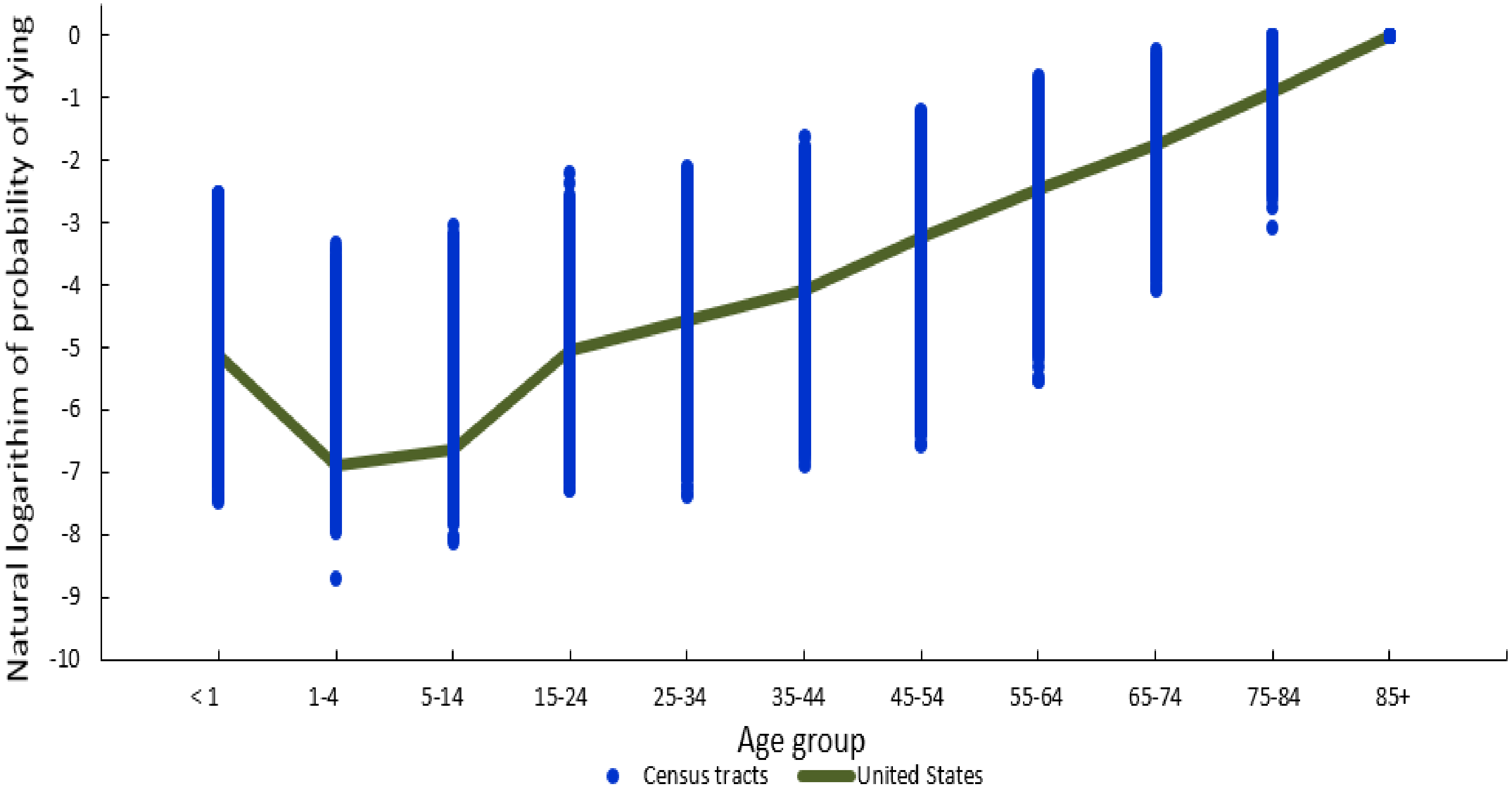
- **Incorporation of population variance into life table functions:**
  - $Var(M_x) = \frac{1}{P^2} * Var(D_x)$
  - $Var(q_x) = \frac{q_x^2(1-q_x)}{D_x}$
  - $Var(e_x) = \sum_x l_x^2 * [(1-a_x)n_x + e_{x+n}]^2 * Var(q_x)$

- ***Delta Method*** $\boldsymbol{Var\big(f(x)\big) = (\frac{df(x)}{dx})^2 * Var(x)}$
  - $\boldsymbol{Var(M_x) = (\frac{dM_x}{dP_x})^2 * Var(P_x) = M_x^2 * (\frac{1}{D} + (\frac{SE(P)}{P})^2)}$
  - $\boldsymbol{Var(q_x) = (\frac{dq_x}{dM_x})^2 * Var(M_x) = [\frac{n_x}{(1+(1-a_x)n_x M_x)^2}]^2 * Var(M_x)}$
  - $\boldsymbol{Var(e_x) = \sum_x l_x^2 * [(1-a_x)n_x + e_{x+n}]^2 * Var(q_x)}$
  - $\boldsymbol{Var(e_{85}) = \frac{l_{85+}^2}{M_{85+}^4} * Var(M_{85+})}$

Final Sample of Census Tract Abridged Life Tables

- After excluding Wisconsin and Maine (will be done separately using 5-years of mortality data),
  - 66,738 tracts had a minimum population of 5,000
  - 65,662 had acceptable life table estimates
  - 8.6% were based exclusively on predicted age-specific death rates

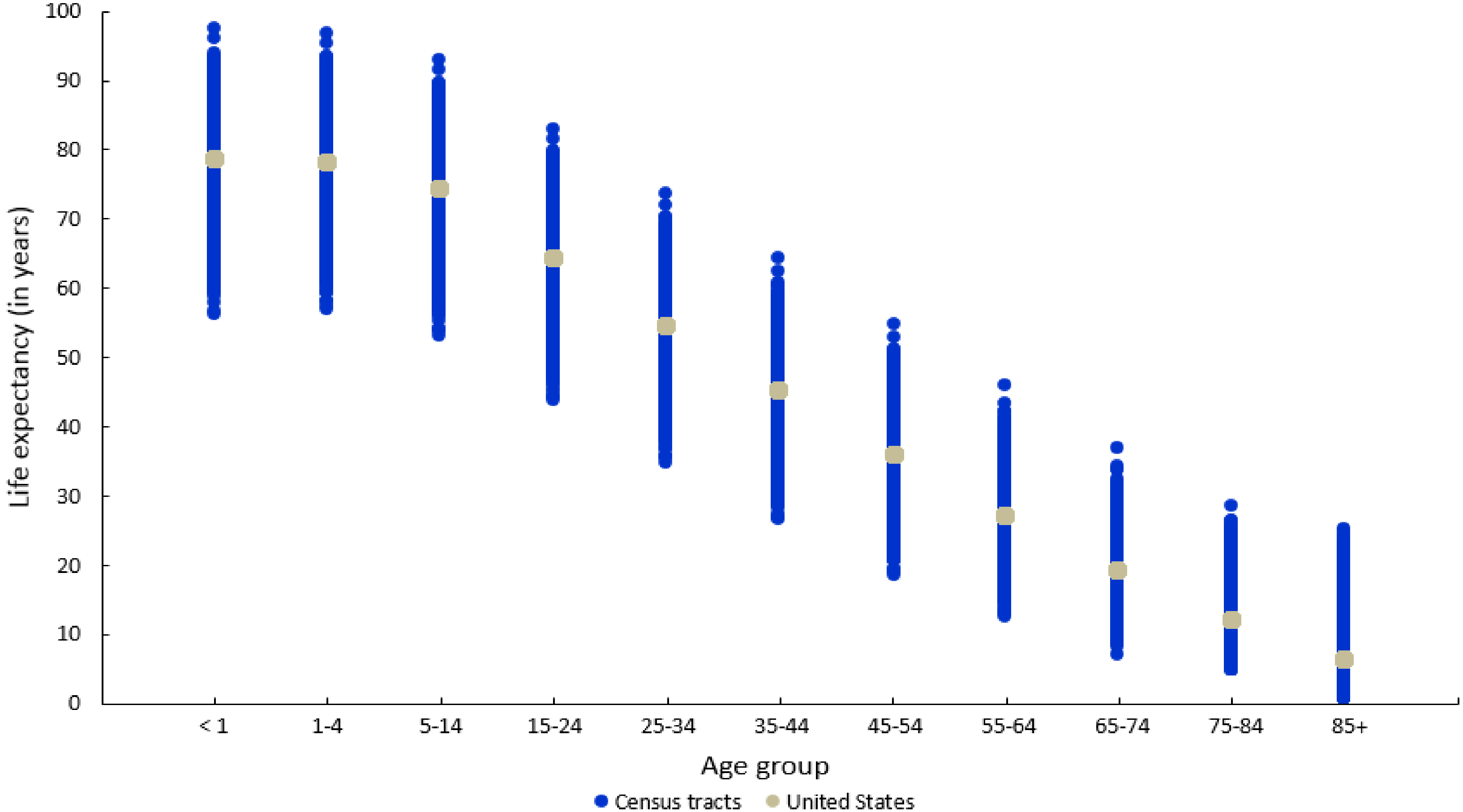
Figure 5. Probability of dying by age in census tracts (2010-2015), compared to United States, 2013



NOTES: Census tracts from Maine and Wisconsin were excluded. Values for the United States, 2013 are based on official life tables: https://www.cdc.gov/nchs/data/nvsr/nvsr66/nvsr66_03.pdf

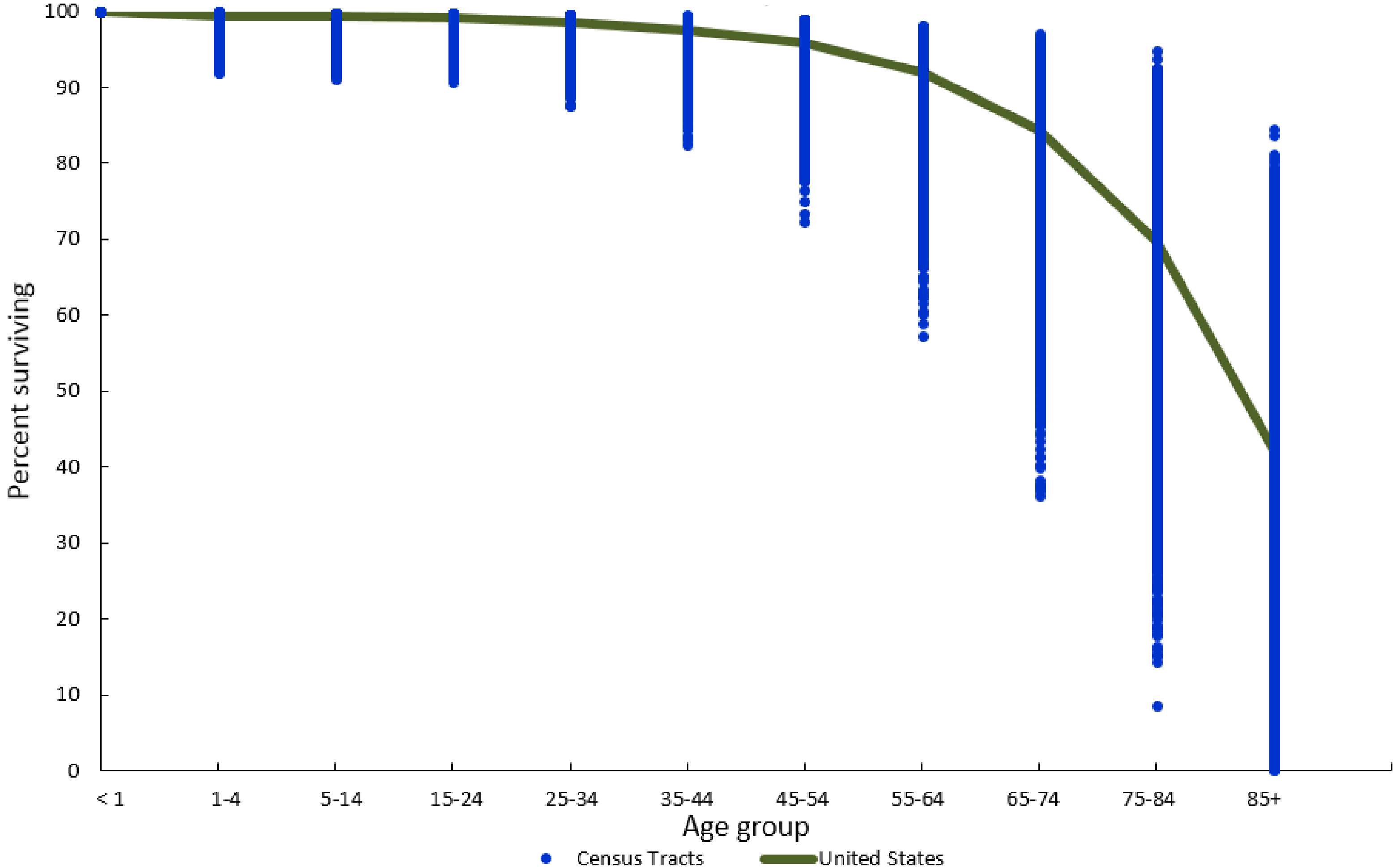
SOURCE: NCHS, National Vital Statistics System, Mortality.

Figure 6. Age-specific life expectancy estimates for census tracts (2010-2015) compared to United States, 2013



NOTES: Census tracts from Maine and Wisconsin were excluded. Values for the United States, 2013 are based on official life tables: https://www.cdc.gov/nchs/data/nvsr/nvsr66/nvsr66_03.pdf
SOURCE: NCHS, National Vital Statistics System, Mortality.

Figure 7. Percent surviving by age in census tracts (2010-2015), compared to United States, 2013



NOTES: Census tracts from Maine and Wisconsin were excluded. Values for the United States, 2013 are based on official life tables: https://www.cdc.gov/nchs/data/nvsr/nvsr66/nvsr66_03.pdf
SOURCE: NCHS, National Vital Statistics System, Mortality.

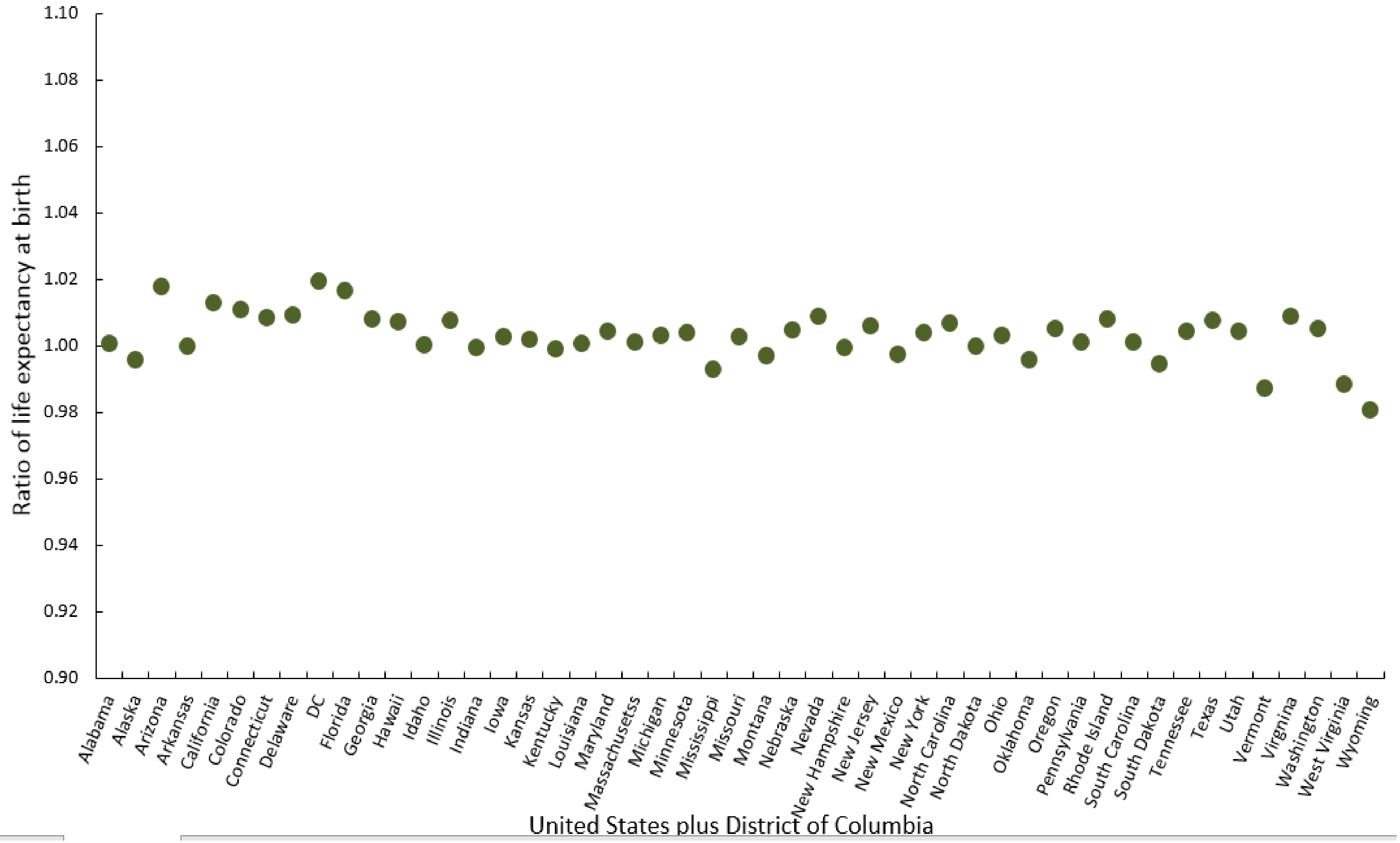
Weighted - Means of Life Table Functions
65,662 Census Tracts, 2010-2015

Age	qx	lx	dx	Lx	Tx	ex	SE ex
0	0.007451	100,000	745	99,329	7,868,455	78.7	1.8021
1-4	0.001520	99,256	151	396,723	7,768,638	78.3	1.9188
5-14	0.001735	99,105	172	990,194	7,371,915	74.4	1.9155
15-24	0.008147	98,934	805	985,308	6,381,721	64.5	1.9135
25-34	0.011898	98,128	1,165	975,457	5,396,412	55.0	1.9015
35-44	0.017813	96,963	1,720	961,035	4,420,954	45.6	1.8903
45-54	0.040411	95,244	3,818	933,345	3,459,919	36.3	1.8908
55-64	0.083300	91,425	7,508	876,712	2,526,574	27.6	1.9233
65-74	0.169839	83,917	13,967	769,337	1,649,862	19.6	2.0261
75-84	0.385691	69,950	26,565	566,677	880,525	12.5	2.2819
85 +	1.000000	43,385	43,385	313,848	313,848	7.1	3.1331

US Abridged Life Table, 2013

Age	qx	lx	dx	Lx	Tx	ex
0	0.005960	100,000	596	99,474	7,882,618	78.8
1-4	0.001016	99,404	101	397,371	7,783,144	78.3
5-14	0.001299	99,303	129	992,435	7,385,773	74.4
15-24	0.006373	99,174	632	989,187	6,393,338	64.5
25-34	0.010574	98,542	1,042	980,476	5,404,151	54.8
35-44	0.016944	97,500	1,652	967,654	4,423,675	45.4
45-54	0.039500	95,848	3,786	942,164	3,456,021	36.1
55-64	0.083846	92,062	7,719	885,908	2,513,857	27.3
65-74	0.172854	84,343	14,579	778,776	1,627,949	19.3
75-84	0.395218	69,764	27,572	572,263	849,173	12.2
85 +	1.000000	42,192	42,192	276,910	276,910	6.6

Figure 8. Ratio of the state weighted means of census tract-level life expectancy at birth to the direct state-level life expectancy at birth estimates, 2010-2015



Summary and Conclusions

- Statistically reliable abridged, period life tables were calculated for a total of 65,662 or 90% of US census tracts (excluding tracts in Maine and Wisconsin) for the period 2010 – 2015.
- The problem of small numbers of deaths, missing age-specific death counts, and small population sizes were addressed through a unique combination of statistical and demographic methods.
- The results of this study concur with the findings of previous research showing that a minimum population size of 5,000 is acceptable but missing death counts cannot be ignored.
- This study has important limitations: lack of decennial census-based population estimates and birth data.

Thank You

USALEEP Production Team:
Loraine Escobedo, PhD
Jocelyn Kennedy, MS
Jodie Cisewski, MS
Chunxia Fu, MS

Brigham Bastian, BS

USALEEP Advisory Committee
Paul Sutton, PhD
Robert Wood Johnson Foundation

Contact Information:
EArias@cdc.gov
301-458-4727